DDI 03424-84

12 JUL 1984

MEMORANDUM FOR: Executive Director

FROM: Robert M. Gates
Deputy Director for Intelligence

SUBJECT: DI Database Development

1. In response to your memo of 4 June, I have given some thought to the relationship between our experiment in Third World databasing and the surge requirement statement in the 1984 addendum to the Strategic Plan. We have, indeed, learned some things in the process of our work that are relevant to the broader issue.

2. DEFINING PURPOSES. Early on in the design of our Third World database, we spent some time trying to sort out its purposes and most likely users. It is intuitively obvious that we ought to build a capability to deal more effectively with spot crises. What is less obvious is whether (1) the database itself should be of a sort that could be used directly by planners engaged in handling the crisis or (2) it should facilitate the sort of analytic support that they will require. (The distinction here is that the planners and executors of emergency programs might well be most interested in local data--such as locations of hospitals, small airfields, or water purification plants--that are very expensive to collect on a large scale and have few applications for our day-to-day work.) As we looked further into this, we concluded that the database we needed--and that we could sensibly maintain--was one of value primarily to our analysts as they worked in support of the policymaker, rather than one that drew every potentially relevant detail on the Third World into one central bank.

3. The construction that we place on this database is that it should be directly useful to analysts and watch officers and--occasionally--transferable to such outside consumers as the White House Crisis Center. This means that it must incorporate our best institutional memory on topics that we know to have been raised over and over again in policymaking on the Third World. By and large, its content would conform to the terms specified in the "1985 Intelligence Capabilities Statement."

4. THE MAINTENANCE LIMITATION. Our own experience and discussions with other agencies that had tried to build large databases made it plain that the principal constraint on the size of the database was bound to be the costs of maintaining it. While large numbers of bits of data could be loaded into a

25X1

ALL PORTIONS
ARE CONFIDENTIAL

SUBJECT: DI Database Development

Third World database, their utility would decline rapidly if we could not periodically validate their accuracy. As the utility of the database declined, its use to us and others--and the incentives to analysts to maintain it--would also decline.

5. We made two early choices to cope with this set of constraints. First, we concluded that the DI's Third World database would, wherever possible, take advantage of quality databases being maintained by other organizations. Second, we pulled out about 30 countries that we felt were sufficiently important to US interests to warrant an extra effort, and agreed that these countries would be the first loaded into the system and that they would always receive more detailed coverage than the others.

6. DESIGN OF A SYSTEM. A problem--and one with implications for further database developments within the Community--is that there is a great range of ADP skills and interests among analysts. Because we felt strongly that the country analysts had to be at the very center of data collection and updating to assure a quality database, we were perforce also committed to a system that would be very user-friendly and would pick up categories that actually captured the key elements of political, economic, and military structure. A fair amount of time has gone into developing three successive prototypes of the database. In one instance, the structure of the database was the product of a country analyst's concerns. In the second and third rounds, this was altered--in response to constructive criticism from other country analysts--to allow greater flexibility for factors important in some countries and not in others.

7. RELEVANT LESSONS LEARNED SO FAR. We believe the modest progress made so far has taught us lessons that are relevant to your broader concerns.

-- Based on extensive conversations with DIA, various civilian agencies, and private-sector concerns, we feel the notion of a high-quality database that can directly service myriad analytic, policy formulation, and operational questions is a dead end. The more purposes the database takes on, the less useful it will be for any one purpose. This is partly related to the risks of design in large systems and partly to the maintenance constraint.

-- The constraint imposed by analysts' ADP capabilities and the sheer software costs of building a comprehensive system under one umbrella mean that a fair share of the content of the database will be in the nature of references to other available databases and off-line sources. This means the database will inevitably have a bit of the directory to it; it will not always bring "the" answer up at the touch of a key.

SUBJECT: DI Database Development

-- We can expect to share materials back and forth between directorates and agencies in tape or disk form, but there will be some costs to adapting and using the other fellow's stuff.

-- Our discussions within the DI with analysts and managers make it plain that construction and maintenance of a Third World database will be expensive in manhours. They also teach us that there is a broad recognition of the need to get on with the job.

-- The process of extracting the data for database construction is, by itself, likely to spotlight information gaps we have only dimly perceived up to this point. A procedure we plan for periodic auditing by functional offices within the DI will likely give us a much better sense of data gaps than we have had in the past. I would expect this to lead to tighter and more timely framing of requirements for our collectors.

8. THE KEY QUESTIONS. Your attached list of questions can be seen in the light of the above comments.

-- If taken as a guiding principle rather than a formula, the definition used in the 1985 capabilites study is still valid. The Third World database we are building now responds to part of the definition, not all of it. It is a combination of text, numbers, and source references that constitute the core of our institutional memory on key Third World counties. This, then, is a subset of all that we have obtained or produced on the Third World--and, therefore, less than the complete database. Moreover, it is not estimative. (DIA chooses to look at this the other way. Its Third World database is comprehensive because it is essentially a catalogue of everything it has produced against its intelligence targets. An issue that has been covered in its database does not necessarily have to be maintained, however.)

-- Clearly, the implication of meeting the terms of this definition within the DI is a program less extensive than the old NIS but more disciplined than what we now do on LDCs. Our analysts will need to have the capacity to both access the Third World database and communicate electronically with other analysts. Beyond this, we should expect to devote more resources to adapting and mounting databases acquired from outside the Agency.

-- The approach to Third World databasing should recognize the specialized interests of components but allow for as much sharing of costs as possible. This means that we should not blunt the analytic

SUBJECT: DI Database Development

usefulness of the DI database by designing it to meet too many needs. On the other hand, information in the DI database should--to the extent feasible--be readily transferrable to systems created by other elements to serve their differing purposes. We should also insist on frequent and extensive communication among members of the Community involved in database development.

-- We believe the Third World database on which we are working must be built to support analysis generally--not just in times of crisis--or there will be too little incentive for the analyst to keep it current.

-- As you can see from what we have learned so far, we believe strongly that the Third World database should capture our institutional memory and enhance our analysis on countries known to be of priority interest to the United States.

-- The scope of the full database will ultimately include all the LDCs shown in the Agency's World Factbook. For those beyond the thirty most important, the detail of data will not substantially exceed the factbook, although source references will tell the analyst where to look for additional data. For the thirty most important LDCs, the OCR subject code sets the boundaries, but not every cell will be filled.

-- The database will be maintained by country analysts and audited by functional analysts in OIA, OGI, and OCR every six months. Resource requirements--in addition to those noted above--will include more slots for both intelligence assistants (two to three per geographic office) and programmers (to maintain and enhance the software).

-- Database development in the private sector demonstrates that the design of successful systems depends on both the inclusion of factors relevant to the user and the application of flexible software. We do not believe that the DI experiment in this area commits the Agency to any particular overarching system in Third World databases. At the same time, we do believe that we are learning things about software, database construction, and maintenance that should prove generally useful to others working in the same area. In short, others building in this field can take what they like and leave the rest.

9. I would be happy to speak to these issues at one of our Friday mornings sessions. Indeed, if it would serve a purpose, we could arrange to demonstrate what we now have in the DI Third World database [redacted] 25X1

Robert M. Gates

[redacted] 25X1

SUBJECT:- DI Database Development

Distribution:
Orig - Addressee
1 - DDI
1 - DDI Registry
1 - Exec Registry
1 - PES Chrono

DI/PES: [redacted] (12June84) 25X1